# The role of co-occurring NPM1 and FLT3 mutations in the acute myeloid leukemia landscape

Timothy Bi

**Introduction**

Acute myeloid leukemia (AML) is a disease characterized by the abnormal proliferation of leukemic blasts due to a differentiation block in the myeloid-lineage pathway of hematopoiesis[1]. This eventually leads to impaired hematopoeiesis and ensuing pancytopenia, which is a common cause both of symptoms (pallor, fatigue, weakness, frequent infections, etc.)[2-3] and of the complications that eventually lead to death[4].

Accurate diagnosis of specific AML subtypes (and thus sensitive treatments) as well as risk classification both depend on the specific chromosomal rearrangements and / or mutations a patient has[5-6]. In particular, there are many recurrent mutations known to drive AML and confer prognosis[7]. Some are general to all cancers, such as TP53, while others are specific to AML. Almost all of these mutations confer prognosis in an independent fashion (i.e. a confirmed mutation in that gene alone is sufficient to assess risk), with two exceptions: NPM1 and FLT3.

NPM1 is a protein that plays an important role in the cell cycle and is involved in ribosome biogenesis, histone chaperoning, centrosome duplication, as well as maintaining the structural integrity of the nucleolus[8]. Its role requires it to effectively shuttle between the nucleoplasm, the nucleolus, and the cytoplasm. FLT3, on the other hand, is a cell-surface receptor tyrosine kinase (RTK). Like most RTKs, it binds to a ligand (FTL3 ligand, or FTL3LG, in this case) and dimerizes, stimulating cellular proliferation and inhibiting apoptosis and differentiation[9]. Constitutively activating mutations result in uncontrolled cell proliferation and independence from FLT3LG binding[10-11].

Current risk assessment criteria based on NPM1 and FLT3 depends on the mutational status and expression levels of both proteins[1,5,7,11]. Surprisingly, mutated NPM1 confers a protective benefit, which can offset the adverse effect of an FLT3 mutation (for instance, mutated NPM1 with high FLT3-ITD is considered intermediate risk, while wild-type NPM1 with high FLT3-ITD is considered adverse risk). The exact molecular mechanism behind the protective benefit of mutated NPM1 in the context of FLT3-activating mutations is currently not well understood[12]. Therefore, this project aims to explore the mutational landscape of NPM1 and FLT3 in AML, and to probe possible shared downstream targets for NPM1 and FLT3 that may explain their coordinated effect on prognosis.

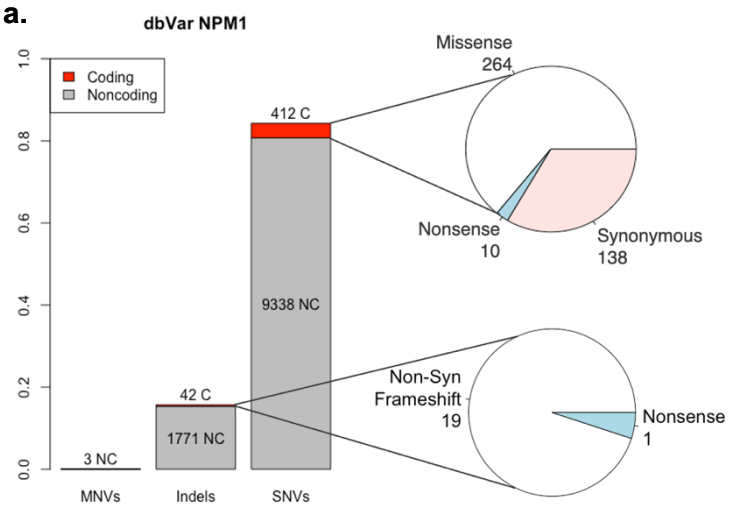
**Results**

*Mutational landscape of NPM1 and FLT3 based on dbSNP.*

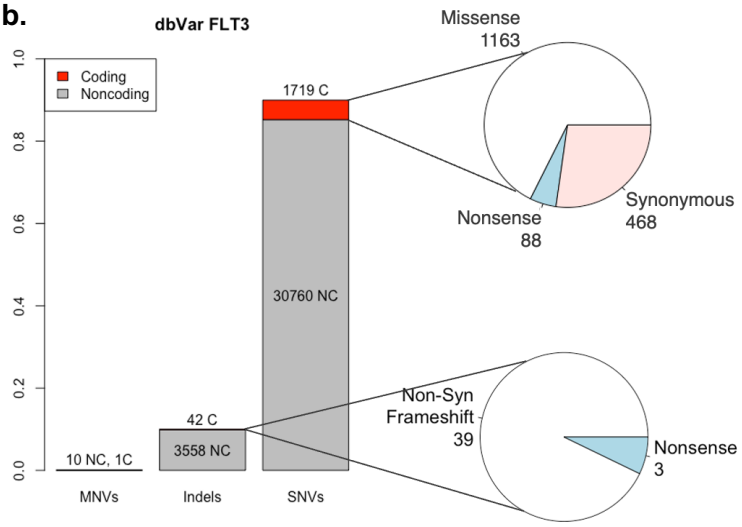
To explore the known mutational landscape of NPM1 and FLT3 and identify common, we chose to analyze known variants catalogued in dbSNP[13], a database maintained by the NIH that catalogs single nucleotide polymorphisms (SNPs) and short deletions and insertions.

For NPM1, we identified 11,566 variants, of which 9750 (84.3%) were single nucleotide variants (SNVs – equivalent to SNPs), 3 (0.026%) were multiple nucleotide variants (MNVs), and 1,813 (15.7%) were insertions / deletions (indels) **(Fig. 1a)**. Of the SNVs, 9,338 (95.7%) variants were non-coding, while 412 (4.2%) variants were coding variants found in exons; of these coding variants, 264 (64.0%) were missense mutations, 138 (33.5%) were synonymous, and 10 (2.4%)

**a.**




**b.**



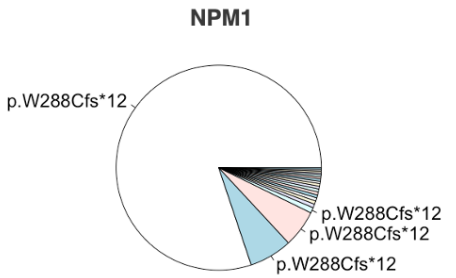

**c.**

| Start Pos | Reference | Alternate | Classification | Variant Type |
|---|---|---|---|---|
| 171410542 | T | TCATG,TCCTG, TTCAG,TTCTG | LP | INDEL, NSF |
| 171410539 | C | CTCTG | P | INDEL, NSF |
| 171410540 | T | TCTGC | P | INDEL, NSF |
| 171410541 | C | CTGCA,CTGCG | P | INDEL, NSF |
| 171410543 | GGCAGTGGAGG | GTTTAAGGATTCG-TC | P | INDEL, NSF |
| 171410548 | TGGAGGAA | TCCCTGGCTAGG | P | INDEL, NSF |
| 171410553 | GA | G | P | INDEL, NSF |
| 171410565 | A | G | P | SNV, SYN |

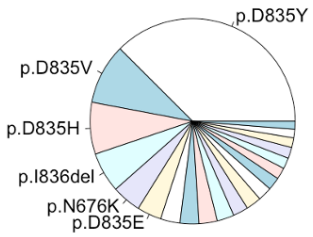
**d.**

| Start Pos | Reference | Alternate | Classification | Variant Type |
|---|---|---|---|---|
| 28018483 | T | A,C | LP | SNV, NSM |
| 28018484 | A | G,T | LP | SNV, NSM |
| 28018485 | G | A,C,T | LP | SNV, NSM, SYN |
| 28018487 | T | A,C,G | LP | SNV, NSM |
| 28018492 | T | A,C,G | LP | SNV, NSM |
| 28018497 | CATG | C | LP | INDEL |
| 28018501 | A | C,G,T | LP | SNV, NSM |
| 28018502 | T | A,C,G | LP | SNV, NSM |
| 28018504 | TC | AA | LP | MNV, NSM |
| 28027222 | A | C,T | LP | SNV, NSM |
| 28027236 | T | A,C | LP | SNV, NSM |
| 28028203 | G | A,C,T | LP | SNV, NSM |
| 28028205 | T | C | LP | SNV, NSM |
| 28028244 | T | G | LP | SNV, NSM |
| 28028279 | T | A,C,G | LP | SNV, NSM |
| 28033974 | C | A,G,T | LP | SNV, NSM |
| 28034139 | A | C,G,T | LP | SNV, NSM |
| 28034144 | A | C,G,T | LP | SNV, NSM |
| 28034147 | T | A,C | LP | SNV, NSM |
| 28034148 | A | C,G,T | LP | SNV, NSM |
| 28034183 | A | C,G | LP | SNV, NSM |
| 28034204 | T | A,C | LP | SNV, NSM |
| 28034205 | A | G,T | LP | SNV, NSM |
| 28018502 | TATC | T | LP/P | INDEL |
| 28018487 | T | TGGATCC | P | INDEL |
| 28018500 | G | A,C,T | P | SNV, NSM, SYN |
| 28018503 | A | C,T | P | SNV, NSM |
| 28034125 | T | TTCATATTCTCT-GAAATCAACG | P | INDEL |
| 28034139 | AATC | A | P | INDEL |

**e.**

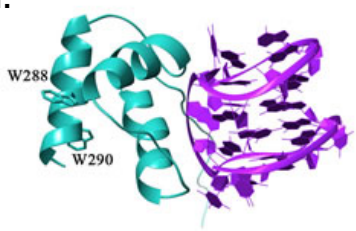



5. p.W288Cfs*12
6. p.?
7. p.W288Cfs*12
8. p.?
9. p.?
10. p.?
11. p.?
12. p.?
13. p.?
14. p.?
15. p.W288Cfs*12
16. p.W288Lfs*12
17. p.W288Sfs*12
18. p.?
19. p.L287Pfs*13
20. p.W290Rfs*10




7. p.A680V
8. p.T227M
9. p.Y597_E598insDYVDFREY
10. p.F594_D600dup
11. p.?
12. p.D835del
13. p.D839G
14. p.D7G
15. p.?
16. p.D96=
17. p.D835E
18. p.V491L
19. p.L576P
20. p.Q989*

**f.**




**g.**

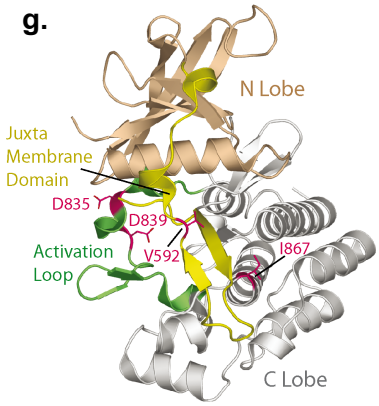



**Figure 1: Analysis of dbSNP, clinVar, and COSMIC for NPM1 and FLT3. a)** Barplot showing all dbSNP records for NPM1 by MNVs, Indels, and SNVs. NC = non-coding, C = coding. Pie charts represent breakdown of types of coding mutations. Same for **b)**, except with FLT3. **c)** and **d)** show dbSNP records with corresponding clinVar annotation of likely pathogenic or pathogenic (shown in classification column) for NPM1 and FLT3, respectively. LP = likely pathogenic, P = pathogenic. SNV = single nucleotide variant, NSM = missense mutation, SYN = synonymous mutation. **e)** Top 20 COSMIC variants for NPM1 and FLT3. Phenotypes with p.? are noncoding intronic variants. **f)** Crystal structure highlighting residues W288 and W290 on NPM1, adapted from Di Matteo et al[14]. **g)** Crystal structure highlighting residues D835 and D839 in the activation loop of FLT3, adapted from Tate et al.[15]

were nonsense. Of the indels, 1,793 (98.9%) are non-coding, and only 20 (1.1%) are coding, with 19 (95%) of the coding variants being non-synonymous frameshift indels and only 1 (5%) nonsense indel. All three MNVs were in non-coding intronic regions.

For FLT3, we identified 36,090 variants, of which 32,479 (90.0%) were SNVs, 11 (0.03%) were MNVs, and 3600 (10.0%) were indels **(Fig. 1b)**. Of the SNVs, 30,760 (94.7%) were non-coding, and 1,719 (5.3%) were coding; of the SNVs in coding regions, 1,163 (67.7%) coded for missense mutations, 468 (27.2%) coded for synonymous mutations, and only 88 (5.1%) coded for nonsense mutations. Of the indels, 3,558 (98.8%) were non-coding, and 42 (1.2%) were coding; of these coding variants, 39 (92.9%) were non-synonymous frameshift mutations, while only 3 (7.1%) were nonsense mutations. Of the 11 MNVs, 10 were non-coding intronic variations, while 1 was a missense mutation.

*Common functional mutations highlighted by clinVar and COSMIC*
Of the 11,566 known mutations in NPM1 and 36,090 known mutations in FLT3, we respectively identified 42 mutations and 151 mutations with corresponding clinical significance annotation in clinVar[14]. Of the 42 NPM1 mutations, 14 were of uncertain significance, 4 were not provided, 16 were benign, 1 was likely pathogenic, and 7 were pathogenic. Likely pathogenic and pathogenic mutations are reported in **Fig 1c**. Notably, four of the eight mutations include an insertion of four nucleobases, which is a frameshift mutation.

Of the 151 FLT3 mutations, 17 were not provided, 77 were benign, 28 were likely benign, 24 were likely pathogenic (1 classified as likely pathogenic / pathogenic), and 5 were pathogenic. As before, likely pathogenic and pathogenic mutations are reported in **Fig 1d**. Of the 24 likely pathogenic mutations, almost all are missense single nucleotide substitutions, and many are in close proximity to the two known pathogenic missense mutations, suggesting a potential hotspot for point mutations in an important functional area of the receptor.

We also used COSMIC[15], an expert-curated database of somatic mutations implicated in cancer, as an orthogonal approach to survey clinically relevant mutations. Because COSMIC documents their mutations by sample rather than by unique mutation, we chose to look at the top 20 mutations by number of samples associated with each mutation **(Fig. 1e)**. The first five most common mutations are all characterized by a substitutions of a tryptophan at position 288 to a cysteine and a resulting frameshift (W288Cfs*12). Notably, this mutation is encoded for by the insertion of a few different variations of a four-nucleotide motif – a finding which matches the pathogenic and likely pathogenic variants we identified in clinVar. And of the 11 protein coding mutations, 9 involve substitution of the tryptophan at position 288 to some other amino acid (and the remaining 2 involve substitutions of nearby residues). All involve an insertion of four nucleobases. The remaining 9 are SNPs in intronic regions.

Applying a similar strategy to FLT3 reveals a much more evenly distributed set of mutations in terms of representation by sample **(Fig. 1e)**. Unlike with NPM1, the top 20 mutations are almost all coding mutations, with only 2 intronic variants. 12 of the coding variants are missense SNPs. We also identify a mutation hotspot around residues 835-839, with residue 835 being the most

commonly mutated. As before, this finding corresponds with what we discovered in clinVar, with most oncogenic mutations being point mutations as opposed to frameshift insertions like in NPM1. These mutations are largely clustered around the activation loop of FLT3 **(Fig. 1g)**.

*Single cell analyses*

To better understand the downstream effects of NPM1 and FLT3 mutations, whether separately or combined, we identified a study containing a cohort of 17 patients with a variety of AML driver mutations[18]. The authors of this paper utilized single-cell genomics to profile bone marrow samples from these patients; specifically to our interests, they identify a patient with neither FLT3 nor NPM1 mutations (AML475), a patient with only a NPM1 mutation (AML556), a patient with only an FLT3 mutation (AML328), and a patient with both (AML210A). They also use a machine learning classifier to classify cells as malignant or benign, obviating the need to use a stem cell score to identify immature blasts.

2D UMAP projection after batch effect correction using Harmony demonstrates the ability to stratify malignant and non-malignant cells **(Fig 2a)**. Cell type annotation was also provided; therefore, we isolated the malignant cells and observe cell type annotations reported in **Fig 2b**. The lack of a malignant lymphoid differentiation trajectory indicates that the differentiation block is only present in the myeloid lineage, which is expected. Furthermore, when we observe the cell type composition of each sample, we find that NPM1-mutated AML is the most differentiated, with a high proportion of monocyte-like and pro-monocyte-like cells **(Fig. 2c)**. AML with neither NPM1 nor FLT3 mutations also tend to have malignant cells that are more mature. FLT3-mutated AML is the most immature, with a high proportion of stem-cell-like and early-progenitor-like cells. Interestingly, AML with both NPM1 and FLT3 mutations appears to be moderately differentiated. Therefore, it appears that NPM1 mutations preserve differentiation capacity, while FLT3 mutations inhibit it. Because earlier differentiation block corresponds to poorer prognosis in AML[19], we hypothesize that the protective effects of NPM1 mutations and the adverse effects of FLT3 mutations are potentially modulated by genes implicated in differentiation.

To identify downstream mediator candidates, we performed differential analysis of malignant cells from each sample and compared each sample to the other. Then, we selected differential genes that were found in every comparison. Some genes of particular functional interest are shown in **Fig 2d**. Specifically, we hone in on a possible TLR4-mediated pathway as a common downstream target for both FLT3-mutated and NPM1-mutated AML. S100A8 and S100A9, which are both highly expressed in NPM1-mutated AML and are expressed at lower levels in FLT3-mutated AML, are implicated in AML differentiation via TLR4[20]. Interestingly, IFI16 is expressed in FLT3-mutated AML but also interacts with TLR4 to drive inflammatory response[21]. Finally, genes that are mediated by the TLR4 signalling cascade, such as JUN, FOS, and ATF4 (whose protein products form a complex known as AP-1), which play a role in cell growth / proliferation, cell differentiation, and programmed cell death, display differential expression between the samples (with generally higher expression in NPM1-mutated AML). This data suggests that the TLR4 signalling pathway may be a possible mechanism by which mutations in NPM1 and FLT3 exert their opposing prognostic effects in AML.

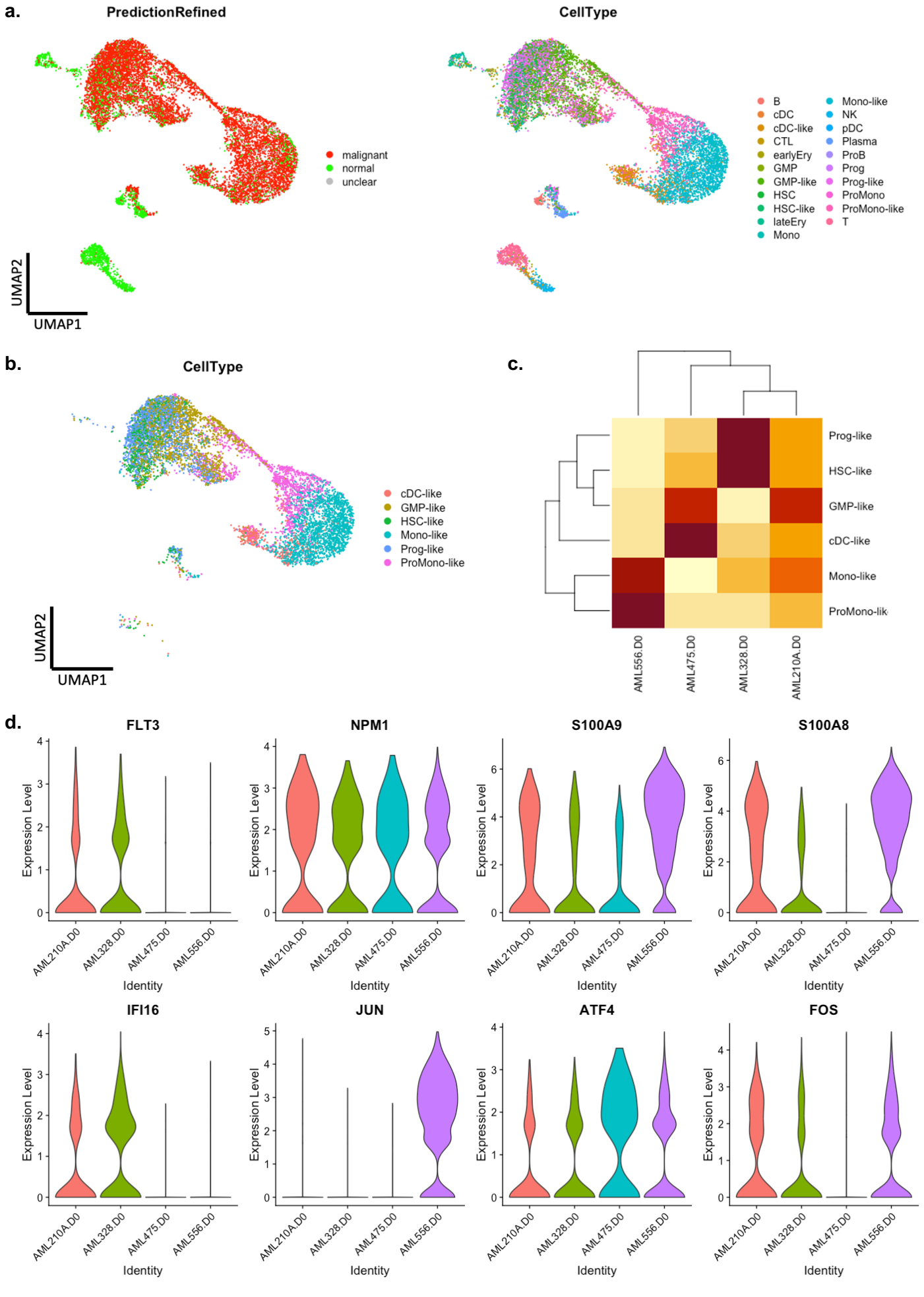

a.
PredictionRefined
malignant
normal
unclear
CellType
B
cDC
cDC-like
CTL
earlyEry
GMP
GMP-like
HSC
HSC-like
lateEry
Mono
Mono-like
NK
pDC
Plasma
ProB
Prog
Prog-like
ProMono
ProMono-like
T
UMAP2
UMAP1
b.
CellType
cDC-like
GMP-like
HSC-like
Mono-like
Prog-like
ProMono-like
UMAP2
UMAP1
c.
Prog-like
HSC-like
GMP-like
cDC-like
Mono-like
ProMono-lik
AML556.D0
AML475.D0
AML328.D0
AML210A.D0
d.
FLT3
NPM1
S100A9
S100A8
IFI16
JUN
ATF4
FOS
Expression Level
Identity
AML210A.D0
AML328.D0
AML475.D0
AML556.D0

**Figure 2: Single-cell analyses of an AML cohort reveal different capacities for differentiation. a)** 2D UMAP projection of scRNA-seq data from bone marrow from 17 patients, colored by malignant status and cell type. **b)** 2D UMAP projection of same data as in **a)**, but with only malignant cells and colored by cell type. **c)** Heatmap showing the percentage of malignant cells in each cell type from four patients AML556 (NPM1m/FLT3wt), AML475 (NPM1wt/FLT3wt), AML328 (NPM1wt/FLT3m), and AML210A(NPM1m/FLT3m). **d)** Differential gene expression for the same patients as in **c)**.

## Discussion

In this report, we attempt to survey the mutational landscape for NPM1 and FLT3, classifying known mutations using dbSNP and annotating functional significance where possible using COSMIC and clinVar. Using a single-cell analysis of bone marrow samples from a cohort containing 17 patients with AML and mutational analysis, we identified that NPM1-mutated AML was more likely to have preserved differentiation capacity, while FLT3-mutated AML demonstrated an earlier differentiation block. Using differential expression analysis between the samples, we were also able to identify the TLR4 signalling pathway as a possible common mediator for the opposing effects on differentiation exhibited by NPM1 and FLT3 mutations.

Our survey of dbSNP for NPM1 and FLT3 mutations display expected frequencies. For instance, we would expect non-coding mutations to be much more frequent than coding mutations, since they are much less likely to affect protein structure / function and thus are not under as much selection pressure. It was surprising, however, that synonymous mutations, which are often assumed to be clinically silent, are less prevalent than nonsense mutations. It may be that these missense mutations occur in locations that are robust to variation and do not cause much structural changes to proteins. On the other hand, the functional impact of synonymous mutations is still being explored[22]. It is also possible that, based on codon usage, there are fewer likely synonymous mutations than functional mutations.

A few major caveats apply to our differential expression analysis. First, the sample size of patients selected is one patient per condition – far too low to draw conclusions with any kind of statistical significance. Second, the presence of other co-occurring mutations was not controlled for – therefore, it is possible that the observed differential expression is due to another mutation, not NPM1 or FLT3. Third, no causal relationship between NPM1 / FLT3 mutations and activation of the TLR4 pathway can be claimed; we only report an association. Finally, the use of computational methods to identify biologically interesting findings must always be followed by validation with *in vitro* and *in vivo* studies to properly define mechanism. This is beyond the scope of our report.

Finally, our final report differs from our proposal in a few ways. We had initially proposed to include dbVar analysis of NPM1 and FLT3 mutations to capture structural variants. However, after some experimentation, we decided to move ahead with only dbVar mutations, along with clinVar and COSMIC annotations for pathogenicity. One downside to this approach is that we were unable to identify FLT3-ITD mutations, which are characterized by recurrent repeats of the FLT gene (which would fall under a category of mutations that is too large to be represented in dbVar). We also did not use a leukemic stem cell (LSC) signature in our single cell analyses, as the dataset that we had chosen already contained annotation of malignant cells. Identifying LSCs by an LSC score and then performing differential expression analyses may have helped us refine our gene selection to better capture LSC self-renewal pathways. Nevertheless, we believe

that this study identifies future directions for further research and a possible therapeutic target in the TLR4 pathway for AML.

**Methods**

*Database preparation and analysis.*

dbSNP VCFs were downloaded from https://ftp.ncbi.nlm.nih.gov/snp/latest_release/VCF/ and preprocessed to identify records specific to NPM1 and FLT3 using the awk command

```
tabix GCF_000001405.40.gz NC_000005.10:171387116-171410900 > NPM1_snvs.tsv
```

for NPM1 and

```
tabix GCF_000001405.40.gz NC_000013.11:28003274-28100576 > FLT3_snvs.tsv
```

for FLT3. All variants were further analyzed in R. The type of variant was identified by the "VC" field in the .vcf file. Coding variants were identified from non-coding variants (and split) using "NSM", "NSN", "NSF", or "SYN" fields. Identification of variants determined to be pathogenic by clinVar used corresponding CLN fields. All code used for this analysis and for all others below is available at ___(insert github here).

COSMIC data was downloaded from https://cancer.sanger.ac.uk/cosmic/download, then filtered for NPM1 and FLT3 using the awk commands

```
awk -F"\t" '{if ($1=="NPM1") {print}}' Cosmic_MutantCensus_v98_GRCh38.tsv >
NPM1_consensus_mutations.tsv

awk -F"\t" '{if ($1=="FLT3") {print}}' Cosmic_MutantCensus_v98_GRCh38.tsv >
FLT3_consensus_mutations.tsv
```

Variants were grouped by their COSV ID, then sorted in descending order for number of samples identifying said mutations. The top 20 mutations were plotted for analysis.

*Single cell analysis*

All scRNA-seq data was downloaded from the Gene Expression Omnibus (Accession ID GSE116256). Read counts and annotations were combined and loaded into a Seurat[23] object. RNA values were transformed using SCTransform[24], then corrected for batch effects using Harmony[25]. 2D UMAPs were generated using the uwot R package[26,27]. Differential gene expression was computed across samples using the FindMarkers() function in Seurat. Code is available at the github referenced above.

**REFERENCES**


1. Pelcovits A, Niroula R. Acute Myeloid Leukemia: A Review. R I Med J (2013). 2020 Apr 1;103(3):38-40. PMID: 32236160.
2. Meyers, C.A., Albitar, M. and Estey, E. (2005), Cognitive impairment, fatigue, and cytokine levels in patients with acute myelogenous leukemia or myelodysplastic syndrome. Cancer, 104: 788-793. https://doi.org/10.1002/cncr.21234
3. Nebgen DR, Rhodes HE, Hartman C, Munsell MF, Lu KH. Abnormal Uterine Bleeding as the Presenting Symptom of Hematologic Cancer. Obstet Gynecol. 2016 Aug;128(2):357-363. doi: 10.1097/AOG.0000000000001529. PMID: 27400009; PMCID: PMC4961605.
4. Klepin HD. Myelodysplastic Syndromes and Acute Myeloid Leukemia in the Elderly. Clin Geriatr Med. 2016 Feb;32(1):155-73. doi: 10.1016/j.cger.2015.08.010. PMID: 26614866; PMCID: PMC4664149.
5. Moarii M, Papaemmanuil E. Classification and risk assessment in AML: integrating cytogenetics and molecular profiling. Hematology Am Soc Hematol Educ Program. 2017 Dec 8;2017(1):37-44. doi: 10.1182/asheducation-2017.1.37. PMID: 29222235; PMCID: PMC6142605.
6. Döhner H, Wei AH, Appelbaum FR, Craddock C, DiNardo CD, Dombret H, Ebert BL, Fenaux P, Godley LA, Hasserjian RP, Larson RA, Levine RL, Miyazaki Y, Niederwieser D, Ossenkoppele G, Röllig C, Sierra J, Stein EM, Tallman MS, Tien HF, Wang J, Wierzbowska A, Löwenberg B. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood. 2022 Sep 22;140(12):1345-1377. doi: 10.1182/blood.2022016867. PMID: 35797463.
7. Kishtagari A, Levine RL, Viny AD. Driver mutations in acute myeloid leukemia. Curr Opin Hematol. 2020 Mar;27(2):49-57. doi: 10.1097/MOH.0000000000000567. PMID: 31972687.
8. Hindley A, Catherwood MA, McMullin MF, Mills KI. Significance of NPM1 Gene Mutations in AML. Int J Mol Sci. 2021 Sep 17;22(18):10040. doi: 10.3390/ijms221810040. PMID: 34576201; PMCID: PMC8467861.
9. Drexler HG, Quentmeier H. FLT3: receptor and ligand. Growth Factors. 2004 Jun;22(2):71-3. doi: 10.1080/08977190410001700989. PMID: 15253381.
10. Kennedy VE, Smith CC. *FLT3* Mutations in Acute Myeloid Leukemia: Key Concepts and Emerging Controversies. Front Oncol. 2020 Dec 23;10:612880. doi: 10.3389/fonc.2020.612880. PMID: 33425766; PMCID: PMC7787101.
11. Kim YK, Kim HN, Lee SR, Ahn JS, Yang DH, Lee JJ, Lee IK, Shin MG, Kim HJ. Prognostic significance of nucleophosmin mutations and FLT3 internal tandem duplication in adult patients with cytogenetically normal acute myeloid leukemia. Korean J Hematol. 2010 Mar;45(1):36-45. doi: 10.5045/kjh.2010.45.1.36. Epub 2010 Mar 31. PMID: 21120161; PMCID: PMC2983002.
12. Sharma N, Liesveld JL. NPM 1 Mutations in AML-The Landscape in 2023. Cancers (Basel). 2023 Feb 12;15(4):1177. doi: 10.3390/cancers15041177. PMID: 36831522; PMCID: PMC9954410.
13. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001 Jan 1;29(1):308-11. doi: 10.1093/nar/29.1.308. PMID: 11125122; PMCID: PMC29783.

14. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018 Jan 4;46(D1):D1062-D1067. doi: 10.1093/nar/gkx1153. PMID: 29165669; PMCID: PMC5753237.
15. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, Fish P, Harsha B, Hathaway C, Jupe SC, Kok CY, Noble K, Ponting L, Ramshaw CC, Rye CE, Speedy HE, Stefancsik R, Thompson SL, Wang S, Ward S, Campbell PJ, Forbes SA. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res. 2019 Jan 8;47(D1):D941-D947. doi: 10.1093/nar/gky1015. PMID: 30371878; PMCID: PMC6323903.
16. Di Matteo, A, Franceschini, M, Chiarella, S, Rocchio, S, Travaglini-Allocatelli, C, & Federici, L. Molecules that target nucleophosmin for cancer treatment: an update. Oncotarget 2016 Jul 12;7(28):44821-44840
17. Janke H, Pastore F, Schumacher D, Herold T, Hopfner KP, Schneider S, Berdel WE, Büchner T, Woermann BJ, Subklewe M, Bohlander SK, Hiddemann W, Spiekermann K, Polzer H. Activating FLT3 mutants show distinct gain-of-function phenotypes in vitro and a characteristic signaling pathway profile associated with prognosis in acute myeloid leukemia. PLoS One. 2014 Mar 7;9(3):e89560. doi: 10.1371/journal.pone.0089560. PMID: 24608088; PMCID: PMC3946485.
18. van Galen P, Hovestadt V, Wadsworth Ii MH, Hughes TK, Griffin GK, Battaglia S, Verga JA, Stephansky J, Pastika TJ, Lombardi Story J, Pinkus GS, Pozdnyakova O, Galinsky I, Stone RM, Graubert TA, Shalek AK, Aster JC, Lane AA, Bernstein BE. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. Cell. 2019 Mar 7;176(6):1265-1281.e24. doi: 10.1016/j.cell.2019.01.031. Epub 2019 Feb 28. PMID: 30827681; PMCID: PMC6515904.
19. Stubbins, R.J., Karsan, A. Differentiation therapy for myeloid malignancies: beyond cytotoxicity. *Blood Cancer J.* **11**, 193 (2021). https://doi.org/10.1038/s41408-021-00584-3
20. Mondet J, Chevalier S, Mossuz P. Pathogenic Roles of S100A8 and S100A9 Proteins in Acute Myeloid and Lymphoid Leukemia: Clinical and Therapeutic Impacts. Molecules. 2021 Mar 2;26(5):1323. doi: 10.3390/molecules26051323. PMID: 33801279; PMCID: PMC7958135.
21. Iannucci A, Caneparo V, Raviola S, Debernardi I, Colangelo D, Miggiano R, Griffante G, Landolfo S, Gariglio M, De Andrea M. Toll-like receptor 4-mediated inflammation triggered by extracellular IFI16 is enhanced by lipopolysaccharide binding. PLoS Pathog. 2020 Sep 9;16(9):e1008811. doi: 10.1371/journal.ppat.1008811. PMID: 32903274; PMCID: PMC7505474.
22. Sauna, Z., Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* **12**, 683–691 (2011). https://doi.org/10.1038/nrg3051
23. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. Integrated analysis of multimodal single-cell data. Cell. 2021 Jun

24;184(13):3573-3587.e29. doi: 10.1016/j.cell.2021.04.048. Epub 2021 May 31. PMID: 34062119; PMCID: PMC8238499.
24. Hafemeister C, Satija R (2019). “Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression.” *Genome Biology*, **20**, 296. doi:10.1186/s13059-019-1874-1
25. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019 Dec;16(12):1289-1296. doi: 10.1038/s41592-019-0619-0. Epub 2019 Nov 18. PMID: 31740819; PMCID: PMC6884693.
26. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018 Dec 3. doi: 10.1038/nbt.4314. Epub ahead of print. PMID: 30531897.
27. Melville J (2023). _uwot: The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction_. R package version 0.1.16, <https://CRAN.R-project.org/package=uwot>.